PHYLOGENETIC STUDIES USING PROTEIN SEQUENCES WITHIN THE ORDER MYRTALES¹

P. G. MARTIN² AND J. M. DOWD²

ABSTRACT

N-terminal sequences (40 amino acids) of ribulose bisphosphate carboxylase small subunit are presented for representatives of Myrtaceae (2 spp.), Melastomataceae (2 spp.), Combretaceae (2 spp.), Lythraceae (2 spp.), Thymelaeaceae (2 spp.), Onagraceae (11 spp.), Buxaceae sensu lato (2 spp.), Euphorbiaceae (2 spp.), and Tiliaceae (2 spp.). In minimal phylogenetic trees the two members of most families pair well, while, in Onagraceae, the phylogenetic tree gives a reasonable approximation to the expectations from taxonomy. The first four families group together as expected for families of Myrtales, but Onagraceae are slightly separate. Thymelaeaceae grouped with the Myrtales families. In order to test the hypothesis that Thymelaeaceae are related to Euphorbiales and both to Malvales, phylogenetic trees embracing all groups were derived. Thymelaeaceae failed to group with Malvales but both showed relationships to Myrtales. Within Euphorbiales and Malvales, grouping of species conformed well with expectations.

A symposium on "The Order Myrtales" was held at the XIII International Botanical Congress in 1981 and, when introducing the published proceedings, Raven (1984) summed up the consensus opinion that this order is a clearly defined group. This suggests that the families that comprise the Myrtales may be favorable subjects for a study in which a comparison is to be made between phylogenies derived from morphological, anatomical, and micromolecular characteristics and from an independent approach such as the computerized analyses of macromolecular sequence data. The most comprehensive attempt to build an angiosperm phylogeny from sequence data is that of Martin et al. (1985), which combined into a single analysis sequences of four proteins and one RNA for up to 11 families. Phylogenetic trees derived from a single macromolecule did not agree with each other and consistency was maintained only when a fourth macromolecule was added to the combined sequences of three proteins. The explanation advanced for this result was that, for four of the macromolecules (the exception was the small subunit of ribulose bisphosphate carboxylase, henceforth RBC-SSU), a family was often (17 out of 33 cases) represented by only a single sequence. Martin et al. (1985) suggested that errors were less likely if a family node was derived from more than one member

of a family. The present investigation was carried out to test this idea that replication is effective, a measure of the correctness of a phylogeny being available in Myrtalean taxonomy. However, in our laboratory it is practicable to sequence only RBC-SSU and there are no sequences of other macromolecules available for the Myrtales. Therefore the test is limited to what is achievable with only one sequence. It can be claimed that, with more sorts of sequence data, greater precision would follow.

MATERIALS AND METHODS

The 27 species that have been investigated are listed by family in Table 1. Methods are set out in detail in Martin and Jennings (1983) and the following is only a very brief account. One hundred grams of fresh leaves were treated using the "pungent-leaf method." After maceration, species of Onagraceae were usually very mucilaginous so that more than the usual amount of extracting buffer was necessary to prevent solidification before gel-chromatography. After purification of the "fraction 1" protein, S-carboxymethylation, and separation of the subunits, five milligrams of small subunit were sequenced on the Beckman Automatic Sequencer 890C. No attempt was made to progress beyond the first 40 amino acids. Amino acids were identified us-

¹ We acknowledge with thanks assistance from the Missouri Botanical Garden and a grant from the Australian Research Grants Scheme; grants to Peter H. Raven from the United States National Science Foundation, most recently BSR82-148, have also supported a portion of this work. We thank Richard Norrish for technical assistance and those who have supplied us with leaves, especially the Missouri Botanical Garden and the Adelaide Botanic Gardens. Peter Raven has given us great encouragement and valuable discussion.

² Department of Botany, University of Adelaide, Box 498, G.P.O., Adelaide, South Australia 5001.

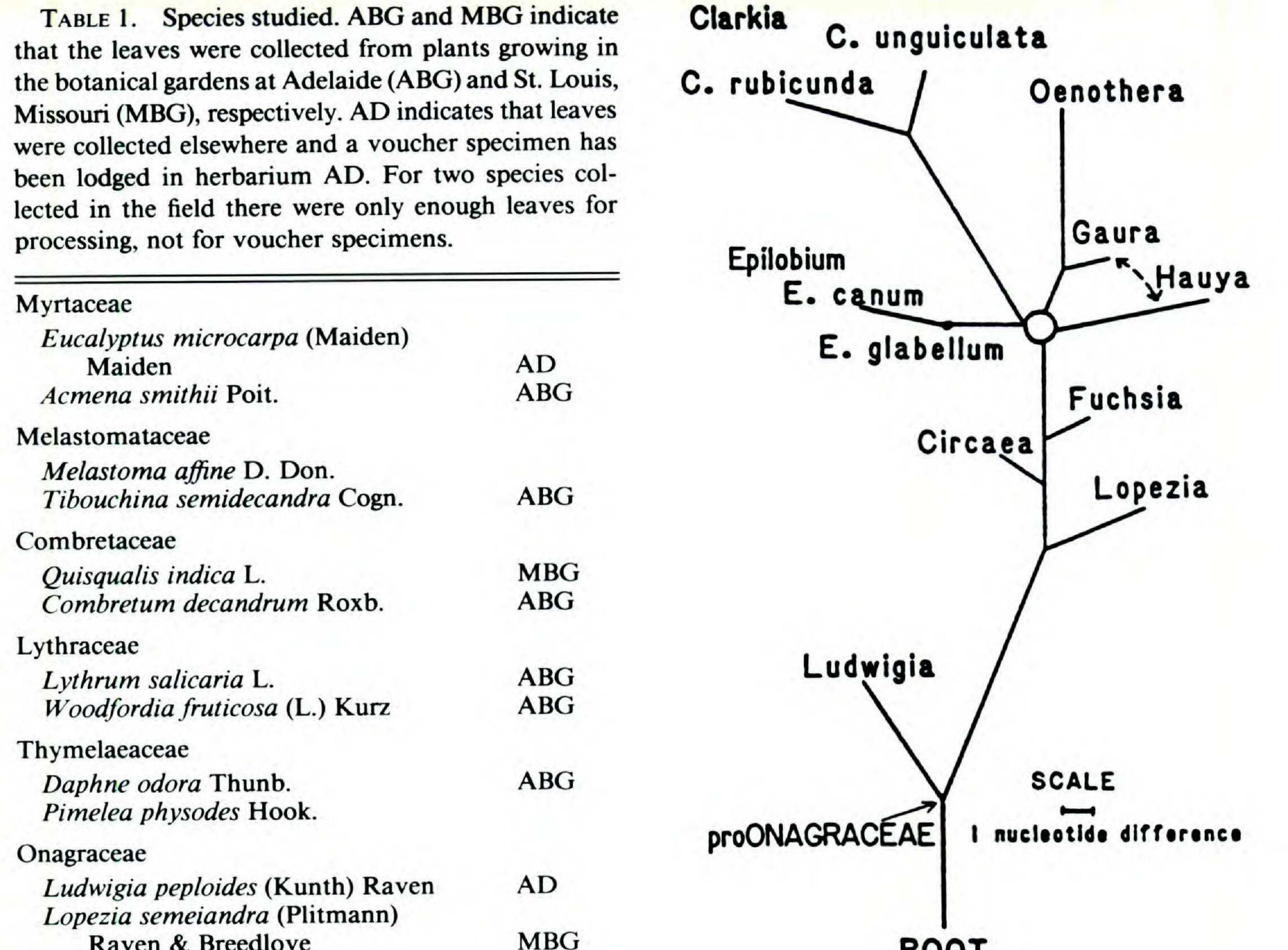
ANN. MISSOURI BOT. GARD. 73: 442-448. 1986.

1986]

MARTIN & DOWD-MYRTALES PHYLOGENY

TABLE 1.

Myrtaceae Eucalyptus microcarpa (Maiden) AD Maiden



MIDO
MBG
AD
AD
ABG
MBG
AD
AD
ABG
AD
ABG
ABG
ABG
ABG

ROOT

443

FIGURE 1. Phylogenetic tree for 11 species of Onagraceae. For specific names see Table 1. The two equally probable consensus trees differed in that the branches carrying Gaura and Hauya were interchanged. Note that in these trees angles are meaningless.

mates the lengths of all possible Steiner trees and the minimal phylogenetic tree is deemed the most probable. Consensus trees are derived from the shortest trees and this procedure is especially important when more than one tree of minimal length occurs. Thus the method is essentially one of maximum parsimony. It should be noted that the number of possible trees increases exponentially with the number of taxa and, with most computers, usually only 11 taxa can be analyzed simultaneously. For taxa that group together constantly no matter what others are present, sequences of their common node can be derived and used in other analyses, so making the treebuilding process more practical.

Grewia occidentalis L.	ABG
Sparmannia africana L.f.	ABG

ing high pressure liquid chromatography and thinlayer chromatography. Sequence data were analyzed by computer methods that have been described by Martin et al. (1985). Briefly the amino acid sequences were converted to inferred nucleotide sequences from which differences between species were derived. The program esti-

RESULTS

General. N-terminal sequences (40 amino acids) are listed in Table 2. More than 100 other

444

ANNALS OF THE MISSOURI BOTANICAL GARDEN

[VOL. 73

TABLE 2. N-terminal sequences of RBC-SSU. For other specific names and also names of families see Table 1. A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Eucalyptus	Μ	Q	v	w	Р	Р	Ι	G	K	K	K	F	Е	Т	L	S	Y	L
Acmena	Μ	Q	v	W	Р	Ρ	Т	G	K	K	K	F	Ε	Т	L	S	Y	L
Melastoma	Μ	Q	V	W	Р	Р	Ε	N	K	K	K	F	Ε	Т	L	S	Y	L
Tibouchina	Μ	Q	V	W	Р	Ρ	Ε	N	K	K	K	F	Ε	Т	L	S	Y	L
Quisqualis	Μ	K	V	W	Р	Ρ	L	G	K	K	K	F	Ε	Т	L	S	Y	L
Combretum	Μ	K	V	W	Р	Ρ	L	G	K	K	K	F	Ε	Т	L	S	Y	L
Lythrum	Μ	Q	V	W	Р	Ρ	Ε	G	L	K	K	F	Ε	Т	L	S	Y	L
Woodfordia	Μ	Q	V	W	Р	Ρ	L	G	K	K	K	F	Ε	Т	L	S	Y	L
Daphne	Μ	Q	V	w	Р	Р	V	G	L	K	K	F	Ε	Т	L	S	Y	L
Pimelea	Μ	Q	v	w	Р	Ρ	Т	N	N	K	1000	F	Ε	Т	L	S	Y	L
Ludwigia	Μ	K	V	w	Р	Ρ	Ε	G	K		100-100	F	Ε	Т	L	S	Y	L
Lopezia	F	Ν	v	W	Р	Ρ	Ε	G	L	L	K	F	Ε	Т	L	S	Y	L
Circaea	F	N	V	W	Р	Р	E	G	L	K	K	F	Ε	Т	L	S	Y	L
Fuchsia	F	Ν	V	w	Р	Ρ	Ε	G	L	K	K	F	Ε	Т	L	S	Y	L
Epilobium ciliatum	F	Ν	v	w	Р	Ρ	Ε	G	L	K	K	F	Ε	Т	L	S	Y	L
Epilobium canum	F	N	V	w	Р	Р	Ε	G	L	K	K	F	E	Т	L	S	Y	L
Hauya	F	Ν	v	W	Р	Ρ	Ε	G	K	K	K	F	Ε	Т	L	S	Y	L
Clarkia unguiculata	F	Ν	v	W	Р	Р	Ε	G	L	K	K	F	Ε	Т	L	S	Y	L
Clarkia rubicunda	F	Ν	V	w	Ρ	Ρ	L	G	L	K	K	F	Ε	Т	L	S	Y	L
Gaura	F	N	V	W	Р	Р	E	G	L	K	K	F	Ε	Т	L	S	Y	L
Oenothera	F	N	v	w	Р	Ρ	Ε	G	L	K	K	F	Ε	Т	L	S	Y	L
Simmondsia	Μ	Q	v	w	Р	Р	L	G	S	K	K	F	Ε	Т	L	S	Y	L
Buxus	Μ	Q	V	W	Р	Ρ	v	G	S	K	Y	F	Ε	Т	L	S	Y	L
Acalypha	Μ	Q	V	W	Р	Ρ	v	G	S	K	Y	F	Ε	Т	L	S	Y	L
Ricinus	Μ	Q	v	W	Р	Р	V/I	G	S		K		Ε	Т	L	S	Y	L
Grewia	Μ	K	V	W	Р	Р	Ι	N	K	K	Ι	F	Ε	Т	L	S	Y	L
Sparmannia		K	V	w	Р	Т	Ε	G	L		K	F	Ε	Т	L	S	Y	L

species of angiosperms have now been analyzed, from a wide range of families, and so it is possible to say that the range of variation exhibited in Table 2 is quite normal in all respects except one. The occurrence of phenylalanine and asparagine at positions one and two in most species of Onagraceae is unknown elsewhere.

Onagraceae. In Raven's (1979) classification of Onagraceae, Zauschneria is combined with Epilobium and Godetia is combined with Clarkia. The results support this taxonomy; preliminary analyses on these sequences showed that Epilobium canum (formerly Zauschneria californica) grouped closely and consistently with Epilobium ciliatum. Similarly Clarkia rubicunda (formerly Godetia amoena) grouped consistently with Clarkia unguiculata. Considering these results, and the taxonomic position, we derived the sequences at generic nodes for Epilobium and Clarkia and used these in subsequent analyses. This reduced the number of taxa from 11 to nine

and for these there were 57 trees of minimal length. This situation arose because there were two adjacent internodes of very short and equal length so that many different combinations were possible without change of overall tree length. For similar reasons there were two consensus trees. We have rationalized the situation and in Figure 1 have presented a single tree with a circle of uncertainty at the relevant node and showing the difference between the two consensus trees. According to Raven (1979) there are seven tribes; Onagreae (represented here by Clarkia, Oenothera, and Gaura), Epilobieae, and Hauyeae arise from this circle of uncertainty. The remaining four taxa in our analysis represent four monogeneric tribes and the analysis shows them as separate. The point at which other families join into this tree (see below) is taken to be the root of the Onagracean tree and this makes the first divergence between Ludwigia and the rest. This is in part because Ludwigia is the only species

1986]

MARTIN & DOWD-MYRTALES PHYLOGENY

TABLE 2. Continued.

19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Р	Т	L	S	S	Ε	Q	L	Α	K	Ε	v	D	Y	L	Ι	K	K	G	w	v	Р
																L					
																N					
																R					
1000	1000			-												•			a second second second second		- Contract of the second se

G K E S W S V D Y L L N S L Q L Q K W N V E D Y K V L S Y E Q L A S L L Ρ G K W E V E Y K V L Ρ S E E L L L P Т L P G W K V Ρ E V Y L K D L L L T Т Q Α P Т P L N K W Y E D L R Ι K Ι E Ι Q A P S L P Т L N K W E Y R Ρ V K E L Ι Ι T Q A S P L P L G K W S V P Y D E K E V L L Q G L P P T D L K N W Y S/R V L P V D E L E K L L Q Q P T P L W G K V Ρ E E R V Y K L L D Q Ι Ρ Т T Q L L G W N V V E S P E Y K E L L Q Ι L Ρ Т Q P L W N G V E V E Y R Ρ L K L E Ι R Q L P P L Т G W N V E E Y R Ρ K V L L R D Q G Ρ P L T L G W N V E Y R Ρ K E V L E L T L Q L P S L P K G W V K Y K E V L Ρ Ι D L Ρ D Q L Ρ Α K N N W V K E Y V D LL Q T D L P P G W E S N E K V T/D Ρ E 0 Р W E N T E V R R Q K Y E P S Ρ G W N Ρ P E S K E E P v A Ρ

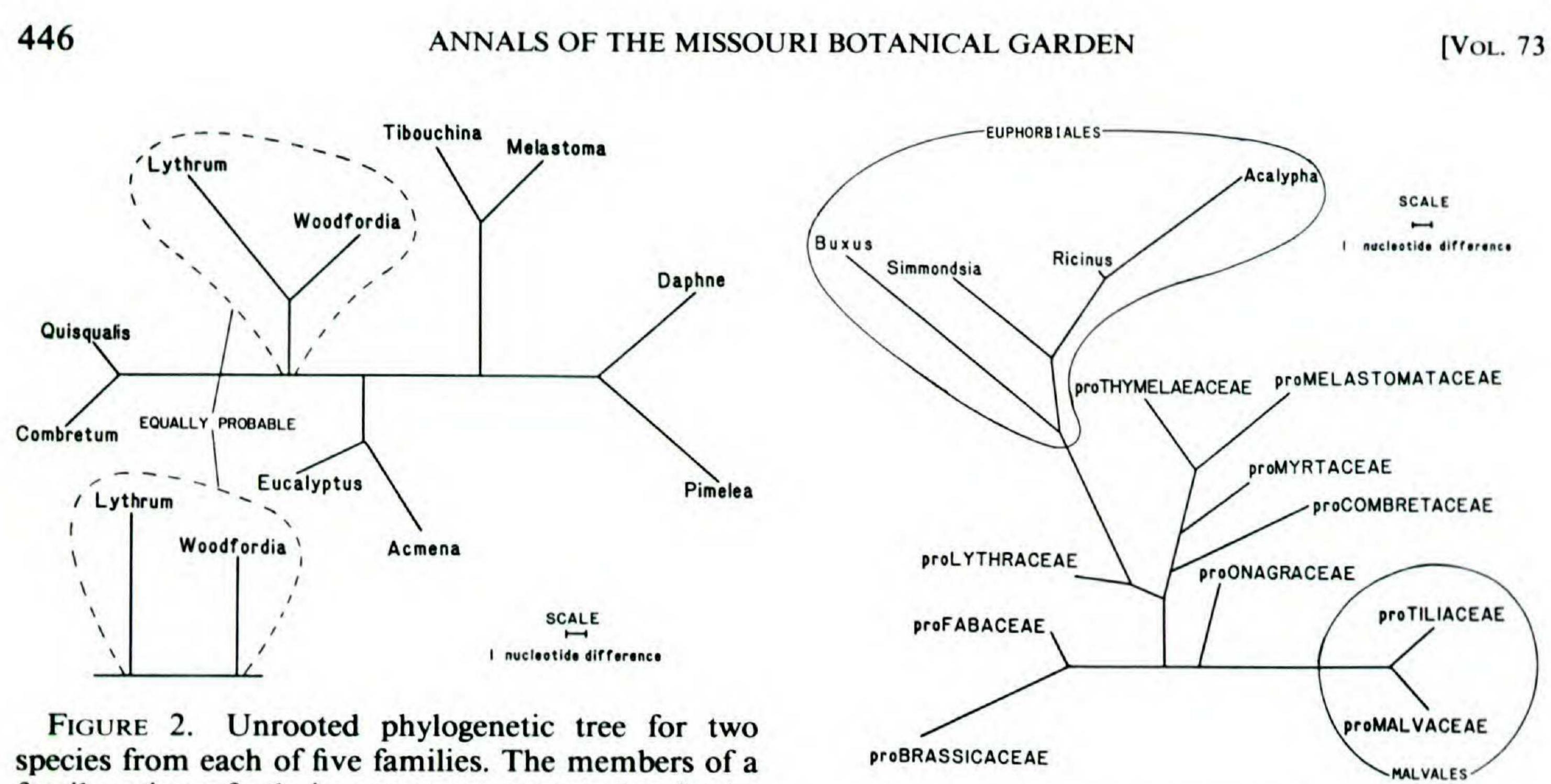
Ρ	Ρ	L	Т	Т	E	Q	L	L	Α	E	Ι	D	Y	L	L	L	S	N	W	V	Ρ
				D																	
				Т																	
				V																	
				D																	

that we have sampled from the family that has a conventional N-terminus, the other ten species having the highly distinctive *Phe-Asn*, which alone accounts for three nucleotide differences.

Four Myrtalean families and Thymelae-In this study four families which, as disaceae. cussed above, are agreed to be members of the order Myrtales, were chosen and, from each, two species, each from a different genus, were selected. Dahlgren and Thorne (1984) considered that, though closely related, Thymelaeaceae was not a Myrtalean family, so it was chosen as an "out-group" (to define the base of the tree) and two species were studied. When these ten species were analyzed simultaneously, there were two consensus trees out of 17 minimal trees (Fig. 2). In one tree the species pair perfectly into the five families, whereas in the other two species of Lythraceae lie side by side. Sequences at the five familial nodes shown in Figure 2 were derived and used in an analysis with Ludwigia, Lopezia,

and Circaea and other external angiosperm nodes (see below) to give the familial node for Onagraceae shown in Figure 1. The grouping of the six familial nodes is shown, with other information, in Figure 3; five of the families arise from a single junction but Onagraceae diverge separately though close to the five. The surprising result in this tree is that Thymelaeaceae appear to be a very definite member of the same group as most of the Myrtalean families and not, as had been expected, an "out-group." Euphorbiales. In one of the currently respected angiosperm taxonomies, Cronquist (1981) did indeed place Thymelaeaceae in Myrtales, although he did so because the family fits in there more easily than anywhere else, not because he considered it a clear-cut Myrtalean family (Cronquist, 1984). Dahlgren and Thorne (1984) stated that a number of taxonomists believe that Thymelaeaceae approach most closely Euphorbiales and they themselves considered that Thymelae-

445



family pair perfectly in one consensus tree but in the equally probable second tree the two members of Lythraceae are side by side. For species names see Table 1. Note that in these trees angles are meaningless. Five of the familial nodes used in deriving Figure 3 were obtained from this tree.

aceae should be placed near Euphorbiaceae, and that these two families are related to Malvales. It therefore seemed appropriate to investigate Euphorbiales and Malvales.

FIGURE 3. Unrooted minimal phylogenetic tree for ten familial nodes and four species from Euphorbiales. The familial nodes were derived from the analyses represented in Figures 1, 2, and 4.

is that Euphorbiales diverge from the Myrtalean tree. We have carried out similar analyses substituting other families for the two "external" ones without changing this result. Thus, except for the junction with the Euphorbiales and Thymelaeaceae, four of the Myrtalean families (Lythraceae, Combretaceae, Melastomataceae, and Myrtaceae) form a natural group as would be expected for an order like Myrtales.

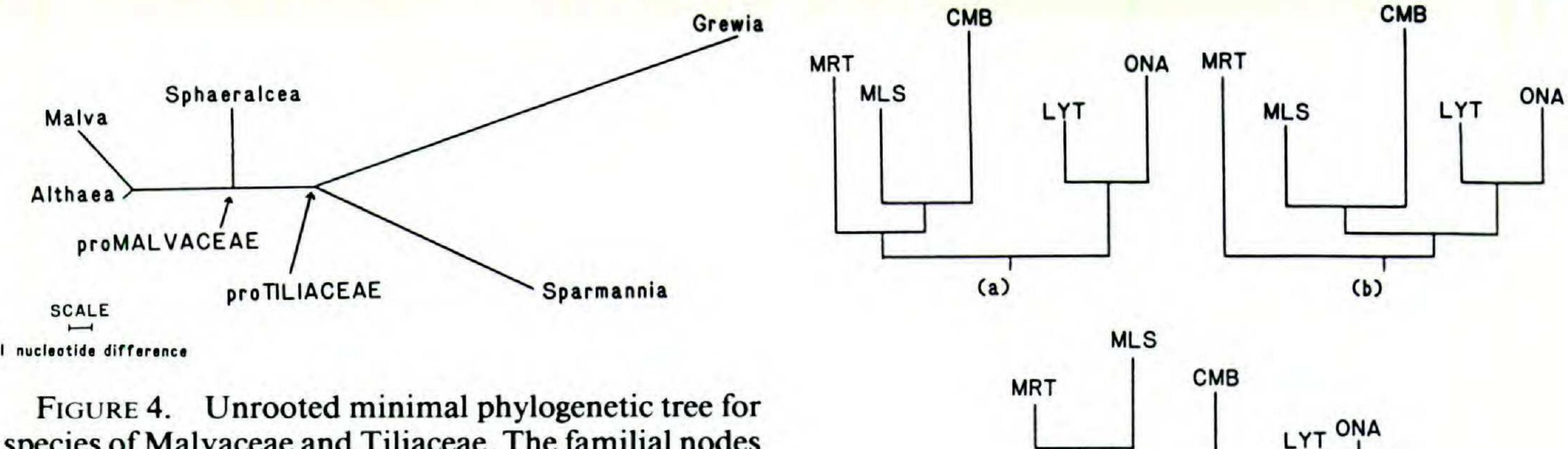
To represent Euphorbiales, two species were chosen from Euphorbiaceae and two from Buxaceae, although it was understood that one of the latter, Simmondsia, is often placed in a family of its own. The four species were first analyzed in conjunction with the Myrtalean familial nodes mentioned above; all four diverged from a common node, confirming the reality of the order Euphorbiales with the two Euphorbiaceae appropriately grouped. However, the two putative Buxaceae did not have a common familial node suggesting that the separation of Buxaceae and Simmondsiaceae might indeed be valid.

In the Malvales, three species of Malvaceae had already been studied (Martin & Dowd, 1984) so, in order to broaden the representation, two species of Tiliaceae were sequenced. These were analyzed with the Malvaceae species and from the minimal tree (Fig. 4) both familial nodes were derived. The analysis of the familial nodes for the five families of Myrtales, Thymelaeaceae, Euphorbiales, and Malvales was carried out with nodes for two other "external" families, Fabaceae and Brassicaceae (Martin et al., 1983) (Fig. 3). Thymelaeaceae do not move from their position in the Myrtales and Onagraceae still diverge slightly separate from most Myrtales. The major surprise

DISCUSSION

As an exercise for testing the ability of objectively-generated phylogenies derived from sequence data to reflect the conclusions of taxonomy, this must be regarded as at least a partial success. Most notable is the correct grouping of species into families as displayed in Figures 2, 3, and 4. In the more extensive study of Onagraceae, two successes can be recorded; the correct grouping of two species each into the genera Clarkia and Epilobium; and the identification of the first divergence, which confirms several other lines of strong evidence that Ludwigia is the "sister group" of all other Onagraceae (Raven, 1979). The separation of the taxa into the other six tribes (Raven, 1979) is more uncertain and three of them (Onagreae, Hauyeae, and Epilobieae) depart from the tree so closely that they are effectively indistinguishable; according to Raven (1979) there is a possible relationship between Hauyeae and Onagreae. The other three tribes (Lopezieae, Circaeeae, and Fuchsieae), which are

MARTIN & DOWD-MYRTALES PHYLOGENY

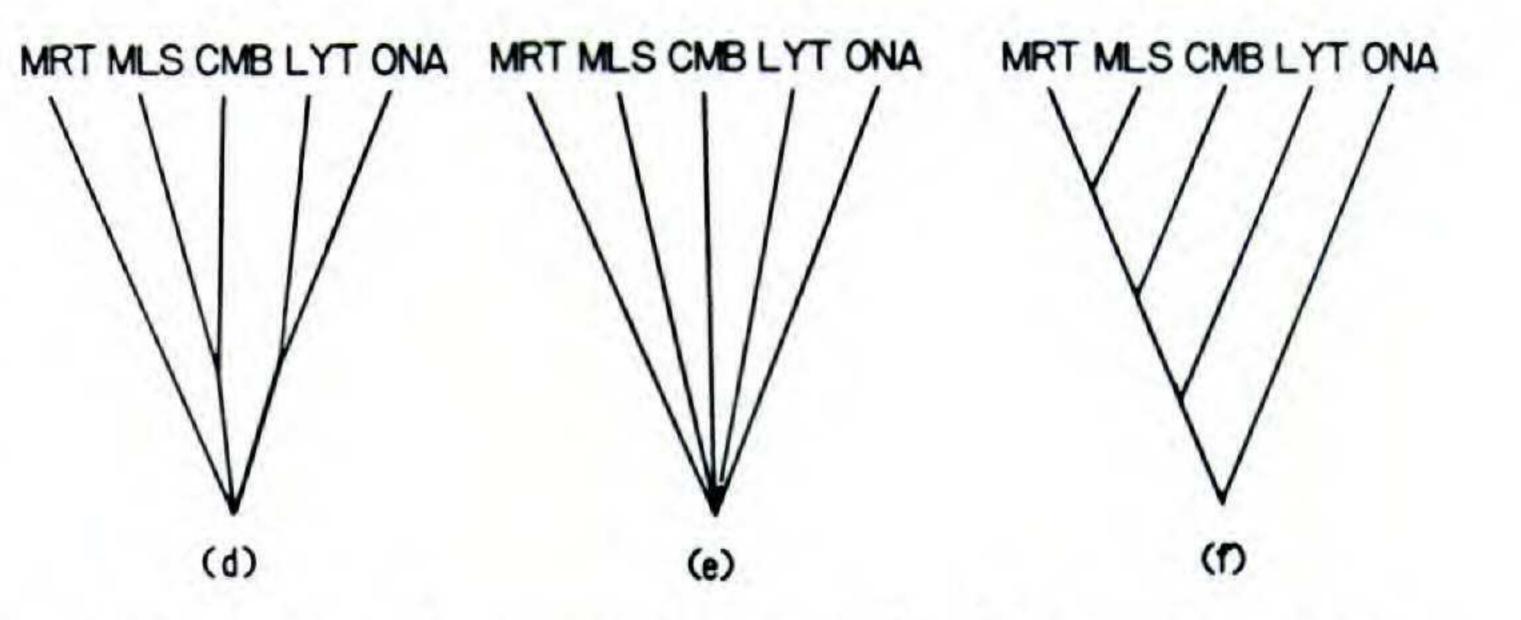


species of Malvaceae and Tiliaceae. The familial nodes for these two families, used in Figure 3, were derived from this tree.

1986]

all monogeneric, are not confused but Fuchsia is thought to be the most generalized and might have been expected to diverge earliest. These are rather stringent tests of our methods because, when relationships are close and differences are small, chance is more likely to operate to give a misleading result; for this reason, Martin et al. (1983) made a strong point of deriving familial nodes from which to generate inter-familial phylogenies, the primary purpose of this research.

When considering the comparison of our phylogenetic trees with those derived for the Myrtales from taxonomy, the best source of comparison appears to be four trees derived from Johnson and Briggs (1984). These authors do not appear to have a clear preference for one or more of these so all are shown in Figure 5; only the families that overlap with this study are represented, viz. Onagraceae, Lythraceae, Combretaceae, Melastomataceae, and Myrtaceae. There appears to be enough variation present among these four trees to accommodate most of the differences in our tree. The main discrepancy is that in some of their trees Myrtaceae diverges near the base; the main similarity is that Onagraceae and Lythraceae usually diverge early. However, it must be said, without in any way implying criticism, that the main objective of the whole exercise has been partly frustrated by the lack of a clearly defined phylogeny to act as a model



(c)

447

FIGURE 5. A comparison of phylogenies from Johnson and Briggs (1984) (a, b, d, and e) with equivalent ones from this paper (c and f). a and b are from Johnson and Briggs' phylograms A and C, respectively, and d and e are congruence cladograms (not to scale) from their figures 6 and 7. MRT, Myrtaceae; MLS, Melastomataceae; CMB, Combretaceae; LYT, Lythraceae; ONA, Onagraceae.

(1984), that of a relationship of Thymelaeaceae to Euphorbiales and Malvales. The results of our test supported Cronquist's hypothesis in that Thymelaeaceae did group with Myrtales, but this result should be considered in the light of the fact that there are sequence data for less than 15% of dicotyledonous families; given tests against a much wider range of variation, the family may well group elsewhere. We believe that the fact that Thymelaeaceae and Euphorbiales do not group with Malvales is stronger evidence against the hypothesis of Dahlgren and Thorne (1984).

In another respect the test has been partly successful because the families of Myrtales do form

against which we can test our methods.

The inclusion of Thymelaeaceae as an "outgroup" for the Myrtales, as suggested by the paper of Dahlgren and Thorne (1984), did not serve that purpose because in our study the family grouped with Myrtales. This was indeed the opinion of Cronquist (1984), so we were able to perform a test, in the best scientific tradition, to distinguish between the hypothesis of Cronquist and the alternative one of Dahlgren and Thorne national Botanical Congress (Raven, 1984).

a natural group separate from others, always excepting Euphorbiales and, possibly, Thymelaeaceae. It is true that Onagraceae seem to diverge separately from the rest but additional data could conceivably change this. It must be acknowledged that Hutchinson (1973) placed Onagraceae in an order of their own but this seems to be quite contrary to the great majority of evidence produced in the symposium at the XIII Inter-

ANNALS OF THE MISSOURI BOTANICAL GARDEN

[VOL. 73

LITERATURE CITED

- CRONQUIST, A. 1981. An Integrated System of Classification of Flowering Plants. Columbia Univ. Press, New York.
- DAHLGREN, R. M. T. & R. F. THORNE. 1984 [1985]. The order Myrtales: circumscription, variation, and relationships. Ann. Missouri Bot. Gard. 71: 633– 699.
- HUTCHINSON, J. 1973. The Families of Flowering Plants, 3rd edition. Clarendon Press, London.

dition of Malvaceae and Ranunculaceae to the phylogenetic tree. Austral. J. Bot. 32: 283-290. — & A. C. JENNINGS. 1983. The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase. I. Biochemical methods and the patterns of variability. Austral. J. Bot. 31: 395-409.

- —, J. M. DOWD & S. J. L. STONE. 1983. The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase.

JOHNSON, L. A. S. & B. G. BRIGGS. 1984 [1985]. Myrtales and Myrtaceae — a phylogenetic analysis. Ann. Missouri Bot. Gard. 71: 700–756.

MARTIN, P. G. & J. M. DOWD. 1984. The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase. III. AdII. The analysis of small subunit data to form phylogenetic trees. Austral. J. Bot. 31: 411-419.
RAVEN, P. H. 1979. A survey of reproductive biology in Onagraceae. New Zealand J. Bot. 17: 575-593.
——. 1984 [1985]. The order Myrtales: a symposium. Ann. Missouri Bot. Gard. 71: 631-632.